Restaurant Sales Data

Most Ordered Product

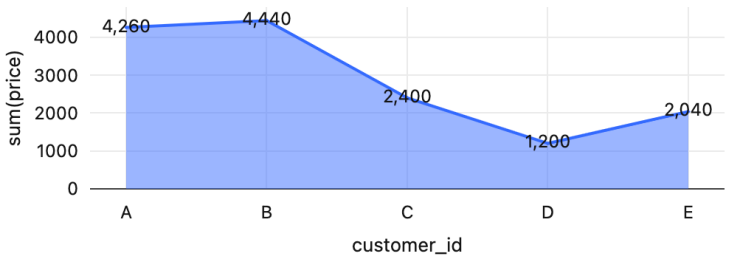
48

(sandwich)

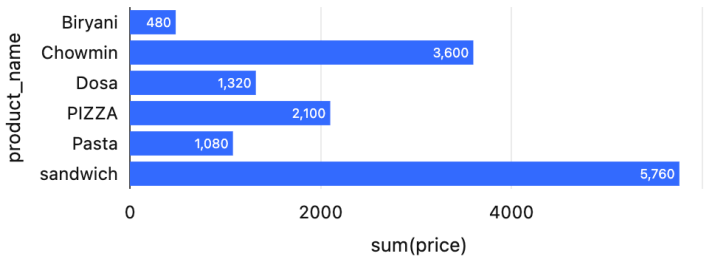
Most Ordered Items

Steps	Value	% Max	% Previous
sandwich	48	100%	100%
Chowmin	24	50%	50%

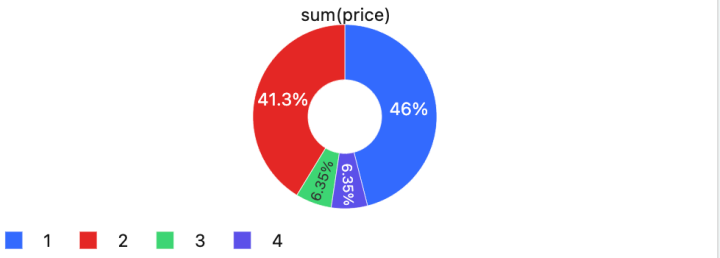
Amount spent by each customer



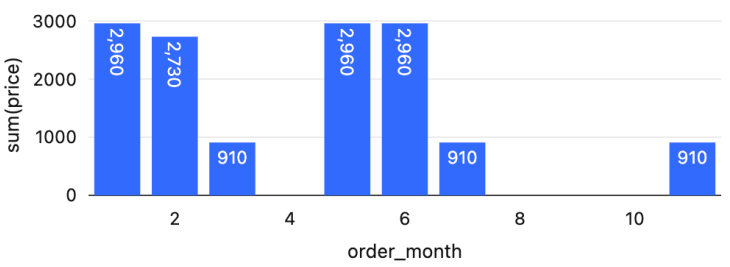
Amount spent by each product category



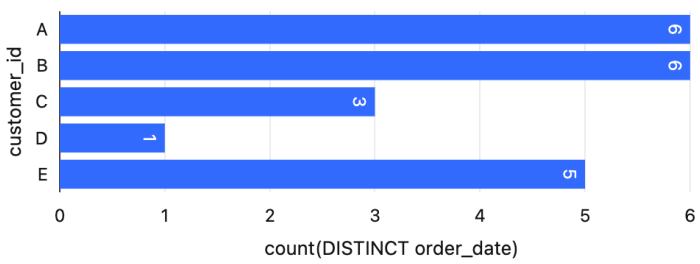
Quarterly Sales



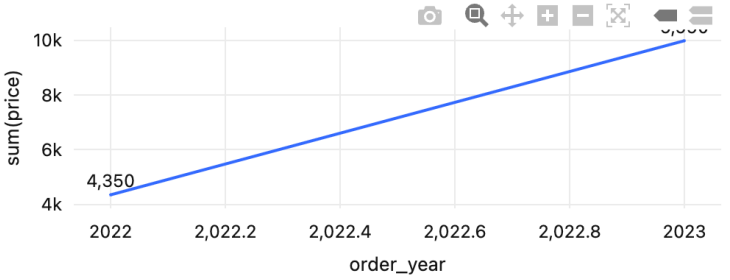
Monthly Sales



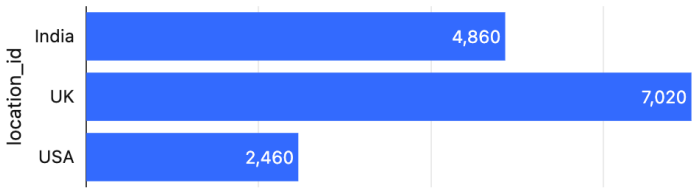
Frequency of order by the customer



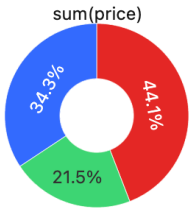
Yearly Sales

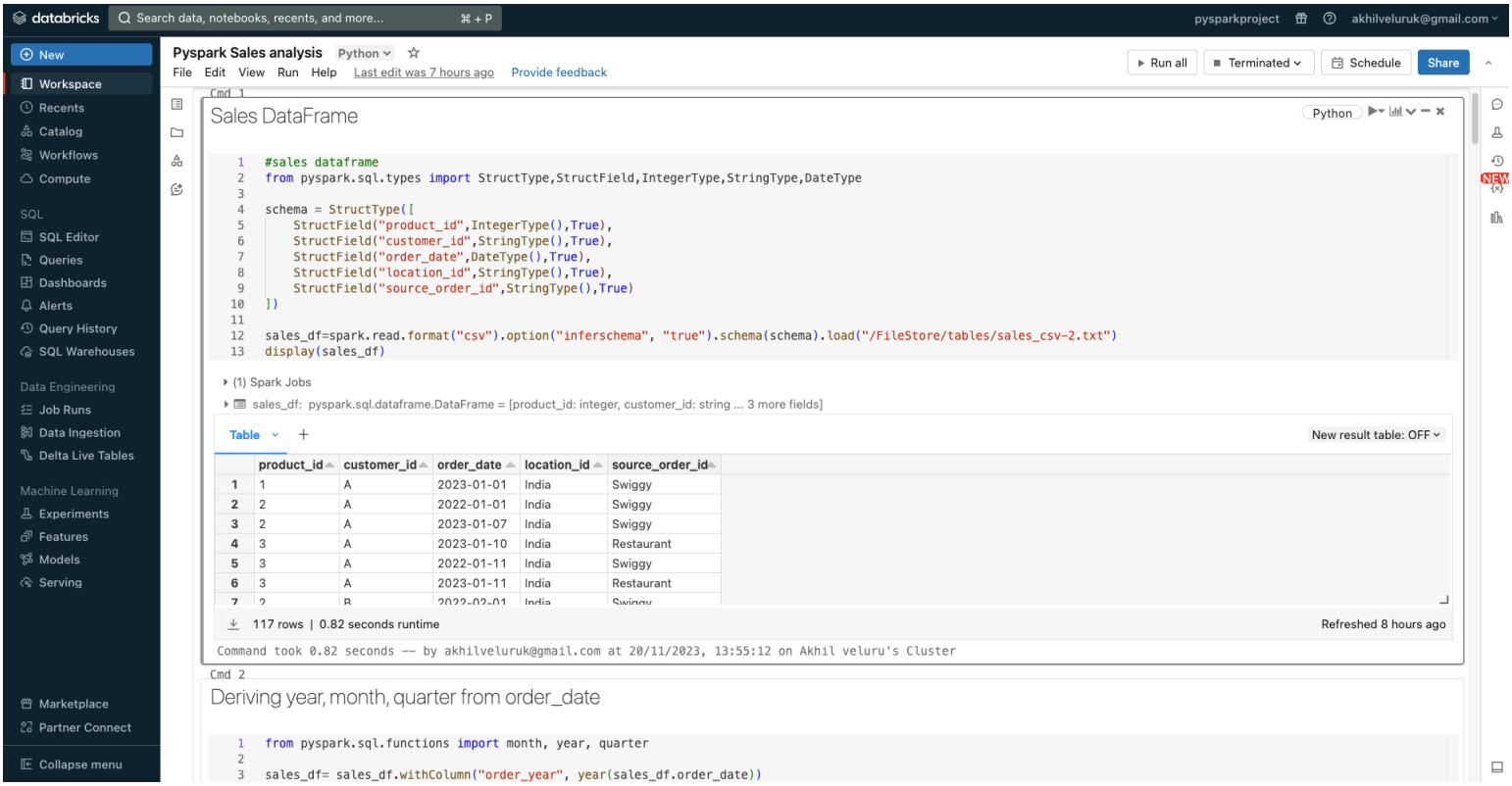


Sales by Country



Order source




databricks
Search data, notebooks, recents, and more... ⌘ + P
pysparkproject
akhilveluruk@gmail.com
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu
Pyspark Sales analysis
Python
File Edit View Run Help
Last edit was 7 hours ago
Provide feedback
Run all
Terminated
Schedule
Share
Cmd 1
Sales DataFrame
Python
#sales dataframe
from pyspark.sql.types import StructType,StructField,IntegerType,StringType,DateType
schema = StructType([
StructField("product_id",IntegerType(),True),
StructField("customer_id",StringType(),True),
StructField("order_date",DateType(),True),
StructField("location_id",StringType(),True),
StructField("source_order_id",StringType(),True)
])
sales_df=spark.read.format("csv").option("inferschema", "true").schema(schema).load("/FileStore/tables/sales_csv-2.txt")
display(sales_df)
(1) Spark Jobs
sales_df: pyspark.sql.dataframe.DataFrame = [product_id: integer, customer_id: string ... 3 more fields]
Table
New result table: OFF
product_id customer_id order_date location_id source_order_id
1 1 A 2023-01-01 India Swiggy
2 2 A 2022-01-01 India Swiggy
3 2 A 2023-01-07 India Swiggy
4 3 A 2023-01-10 India Restaurant
5 3 A 2022-01-11 India Swiggy
6 3 A 2023-01-11 India Restaurant
117 rows | 0.82 seconds runtime
Refreshed 8 hours ago
Command took 0.82 seconds -- by akhilveluruk@gmail.com at 20/11/2023, 13:55:12 on Akhil veluru's Cluster
Cmd 2
Deriving year, month, quarter from order_date
from pyspark.sql.functions import month, year, quarter
sales_df= sales_df.withColumn("order_year", year(sales_df.order_date))